# A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome

**Resource**

Ulrich Stelzl,[1] Uwe Worm,[1] Maciej Lalowski,[1]
Christian Haenig,[1] Felix H. Brembeck,[1]
Heike Goehler,[1] Martin Stroedicke,[1]
Martina Zenkner,[1] Anke Schoenherr,[1]
Susanne Koeppen,[2] Jan Timm,[1] Sascha Mintzlaff,[1]
Claudia Abraham,[1] Nicole Bock,[2] Silvia Kietzmann,[2]
Astrid Goedde,[3] Engin Toksöz,[1] Anja Droege,[1]
Sylvia Krobitsch,[2] Bernhard Korn,[3]
Walter Birchmeier,[1] Hans Lehrach,[2]
and Erich E. Wanker[1,2,*]
[1]Max Delbrueck Center for Molecular Medicine
13092 Berlin-Buch
Germany
[2]Max-Planck-Institute for Molecular Genetics
14195 Berlin
Germany
[3]German Genome Resource Center
69120 Heidelberg
Germany

## Summary

**Protein-protein interaction maps provide a valuable framework for a better understanding of the functional organization of the proteome. To detect interacting pairs of human proteins systematically, a protein matrix of 4456 baits and 5632 preys was screened by automated yeast two-hybrid (Y2H) interaction mating. We identified 3186 mostly novel interactions among 1705 proteins, resulting in a large, highly connected network. Independent pull-down and coimmunoprecipitation assays validated the overall quality of the Y2H interactions. Using topological and GO criteria, a scoring system was developed to define 911 high-confidence interactions among 401 proteins. Furthermore, the network was searched for interactions linking uncharacterized gene products and human disease proteins to regulatory cellular pathways. Two novel Axin-1 interactions were validated experimentally, characterizing ANP32A and CRMP1 as modulators of Wnt signaling. Systematic human protein interaction screens can lead to a more comprehensive understanding of protein function and cellular processes.**

## Introduction

Protein-protein interactions (PPIs) are crucial for all biological processes. Therefore, compiling PPI networks provides many new insights into protein function. Also, interaction networks are relevant from a systems biology point of view, as they may help to uncover the generic organization principles of functional cellular networks, when both spatial and temporal aspects of interactions are considered (Ge et al., 2003). The gener-

ation of accurate cellular protein interaction networks is an ongoing process, in which data produced by high-throughput yeast two-hybrid screens (Y2H) and mass spectroscopy contribute in a complementary manner (Bork et al., 2004).

The Y2H system is a powerful tool for the identification of PPIs, which can be applied in high-throughput manner to detect interactions across the entire proteome of an organism. Proteome-wide studies for model organisms such as *H. pylori* (Rain et al., 2001), *S. cerevisiae* (Ito et al., 2001; Uetz et al., 2000), *C. elegans* (Li et al., 2004), and *D. melanogaster* (Giot et al., 2003) have been performed, yielding large Y2H interaction maps, which are currently utilized for more detailed experimentation and formulation of biological hypotheses. However, a proteome-wide interaction map of human proteins has not yet been provided. Taking into account the high potential of human PPIs for understanding disease mechanisms and signaling cascades, smaller scale, more specific interaction maps have been generated. Examples are an interaction network for Huntington's disease with 186 interactions (Goehler et al., 2004) or a network for the transforming growth factor-β (TGF-β) signaling pathway with 755 interactions (Colland et al., 2004). In addition, bioinformatic analyses have been performed, collecting information on human interactions from hypothesis-driven studies (Peri et al., 2003) or from studies identifying conserved orthologous interactions (Lehner and Fraser, 2004), which are referred to as "interologs" (Matthews et al., 2001; Walhout et al., 2000). Nevertheless, the transfer of interaction information from model organisms to human is far more difficult than anticipated (Bork et al., 2004; Ramani et al., 2005). Therefore, systematic mapping of human protein interactions is indispensable for annotating human protein function and understanding complex cellular processes.

The present screening was set up in order to generate a PPI map that describes a representative part of the human interactome. In contrast to the high throughput library screens used to analyze model organisms (Formstecher et al., 2005; Giot et al., 2003; Li et al., 2004), we have carried out a matrix interaction mating screen. This approach yields reproducible interaction data without the necessity of repeated sequencing, once a bait/prey protein matrix is established. We have systematically screened more than 5500 human proteins for potential interactions, building an interaction network that connects 1705 human proteins via 3186 interactions. The quality of the dataset was validated by independent biochemical interaction assays and bioinformatic analyses. Furthermore, we have evaluated the network to identify proteins potentially involved in human regulatory pathways and have characterized experimentally two novel modulators of the Wnt signaling cascade. The network presented here can be regarded as a starting point for the construction of a
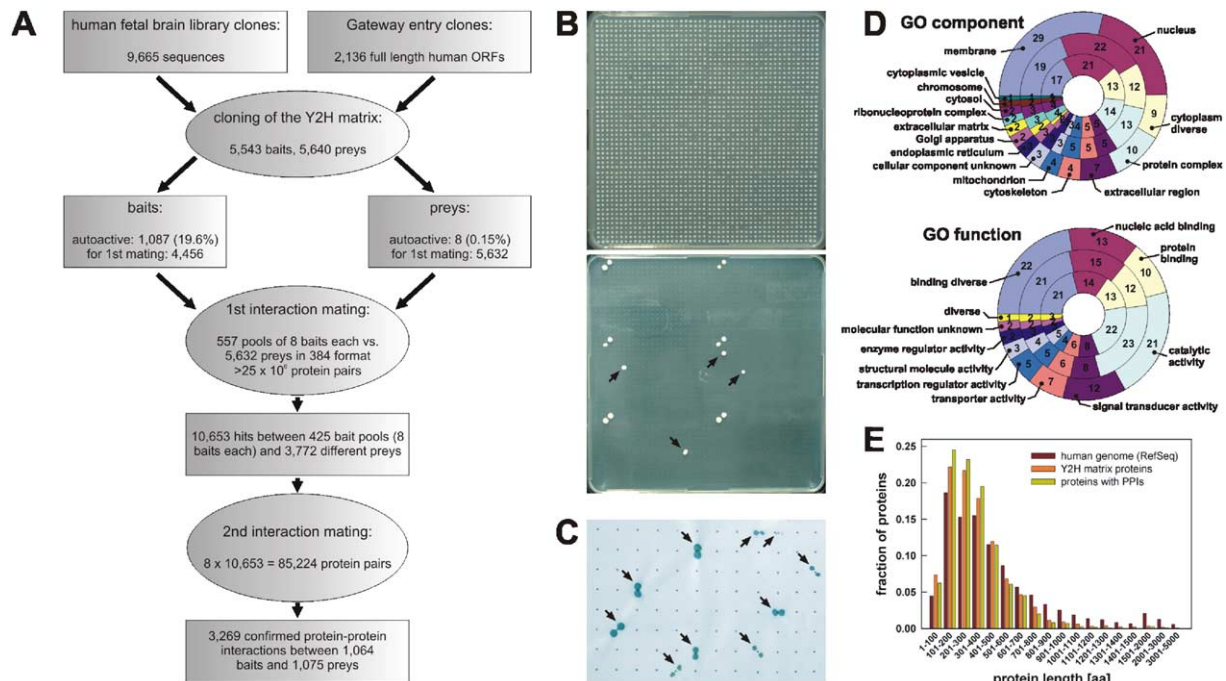
*Correspondence: erich.w@mdc-berlin.de

Figure 1. Automated Y2H Matrix Interaction Mating

(A) Process of the systematic automated large-scale Y2H matrix interaction mating. Three major steps: (1) creation of a human protein matrix in yeast, (2) high-throughput screening of the Y2H matrix with pools of eight baits (first interaction mating), and (3) confirmation of identified interactions using individual pairs of baits and preys (second interaction mating).

(B) Identification of Y2H PPIs using the pooled mating approach (first interaction mating). Top panel: yeast clones spotted onto SDII (-Trp-Leu) agar plates to select for diploid yeasts expressing bait and prey fusions. Bottom panel: PPIs identified (arrows) by assaying the growth of diploid yeasts on SDIV (-Trp-Leu-His-Ura) agar plates.

(C) Confirmation of Y2H interactions by analyzing yeast clones expressing single pairs of bait and prey proteins (second interaction mating). Diploid yeast clones were gridded in duplicates onto nylon membranes (3 × 3 pattern) placed on SDIV agar plates. Membranes were assayed for β-galactosidase activity to identify positive clones (arrows).

(D) Distribution of cellular component and molecular function GO categories. Outer rings: proteins encoded by the human genome (10,504 with GO component and 12,174 with GO function identifiers). Middle rings: proteins contained in the Y2H matrix (3267 with GO component and 3778 with GO function identifiers). Inner rings: PPI network proteins (1064 with GO component and 1208 with GO function identifiers). Each section represents the number of proteins (percentage indicated) assigned to a given GO category. The Y2H matrix is a representative subset of all human proteins. The distribution between categories did not change for network proteins.

(E) Length distribution of human proteins. Red bars: fraction of proteins encoded by the human genome (28,707 according to RefSeq [NCBI]). Orange bars: proteins contained in the Y2H matrix (5640). Yellow bars: proteins with Y2H interactions (1824). The distribution of the matrix proteins was similar to the proteins encoded by the genome. The majority of the proteins (81% of the matrix proteins and 70% of the proteins in the human genome) had a predicted length between 100 and 500 amino acids.

more comprehensive human PPI map and provides a useful resource for further elucidation of protein function.

## Results

### Construction of a Human Y2H PPI Map

For the identification of interactions by a Y2H matrix approach (Figure 1A), a nonredundant set of cDNAs was obtained from the sequence analysis of a human fetal brain expression library (Bussow et al., 1998). We generated 3510 bait and 3589 prey clones by subcloning of cDNA fragments into DNA binding domain (DBD) and activation domain (AD) Y2H vectors, respectively. In addition, a "GATEWAY recombinational cloning" approach was used to shuttle full-length human open reading frames (ORFs) from entry vectors into Y2H plasmids, yielding further 2033 bait and 2051 prey clones.

From the total of 11,183 Y2H clones, we created a matrix for systematic interaction mating. A *MATα* yeast strain was individually transformed with the 5640 prey plasmids, and 5543 bait plasmids were introduced into a *MAT*a strain. For interaction screening, eight clones expressing non-self-activating baits were pooled and mated with the prey clones. Positive clones, which activated the *HIS3* and *URA3* reporter, were identified by growth on selective plates (Figure 1B) and/or *lacZ* reporter gene activation in β-galactosidase assays. All positive preys identified in the first pooled mating screens were individually retested for interactions with each of the eight baits in a second mating assay (Figure 1C). An unambiguously positive interaction was only assigned to a pair of proteins after two independent mating assays.

More than 25 million protein pairs (4456 baits × 5632 preys) were examined, and 3269 interactions among 1064 baits and 1075 preys were identified (Figure 1A

and see Tables S1–S3 in the Supplemental Data available with this article online). The Y2H interactions were grouped into two data sets: those that activated all three reporter genes, *HIS3*, *URA3*, and *lacZ* (LacZ4 set), and those that activated only the two growth reporters *HIS3* and *URA3* (SD4 set). The LacZ4 and SD4 sets comprise 2124 and 1145 interactions, respectively.

To examine whether the protein matrix is an unbiased representation of the human proteome, gene ontology (GO) criteria were applied (Ashburner et al., 2000). Except for membrane proteins, which were underrepresented because the library used for subcloning was selected for cDNAs that encode proteins without transmembrane domains, similar size fractions of proteins annotated to the different GO component as well as GO function categories (Figure 1D, upper and lower panel). Interestingly, the distribution of proteins in both the GO function and component classes remained largely the same among the identified interactions as in the matrix and the complete proteome (Figure 1D, inner circles). In Figure 1E, the protein length distribution of the matrix and the interacting proteins was compared with all human ORFs encoded in the genome. In all three groups, the majority of proteins had a predicted length between 100 and 500 amino acids.

## Experimental Verification of Interactions

To evaluate the quality of the Y2H data, a representative sample of interactions was randomly selected for verification assays because interactions recapitulated independently are unlikely to be experimental false positives (Goehler et al., 2004). Using a membrane coimmunoprecipitation assay (Figure 2A), 116 protein pairs were tested with a success rate of 72/116 (62%). Another 131 interactions were verified in pull-down experiments (Figure 2B), with a success rate of 87/131 (66%). With both assays, a difference in success rates was observed when the SD4 (56%) and the LacZ4 data (67%) were compared (chi-square p = 0.11). These results demonstrate that our Y2H data contain a large fraction of interactions that can be confirmed by other methods. They also support previous Y2H studies (Vidalain et al., 2004), emphasizing the difference in validity of the interactions from the SD4 and the LacZ4 sets.

## Properties of the Y2H Network

For analysis of the protein interaction data, each cDNA was mapped to an NCBI gene locus (Table S2). After collapsing 83 interactions that occurred in bait/prey and prey/bait configurations or that mapped pairwise to the same gene loci, a total of 3186 unique interactions between 1705 different human proteins was obtained (Tables S1 and S3). Computational analysis of the data revealed one giant network of 3131 interactions between 1613 proteins and 43 small isolated networks of less than six proteins (Figure S2). For the large interaction network, a mean shortest path length between any two proteins of 4.85 links was calculated (Figure 3A). This means that most proteins are very closely linked, a phenomenon that has been described as small world property of networks (Strogatz, 2001).

Next, we calculated the degree distribution P(k) of the human proteins, measuring the probability that a given protein interacts with k other proteins. As shown in Figure 3B, the degree distribution of the network proteins decreases slowly, closely following a power-law. This indicates that the human interaction map has scale-free properties (Barabasi and Oltvai, 2004), which is in agreement with interaction studies for model organisms (Figure S3A). On average, proteins in the network have 1.87 interaction partners. However, 804 proteins with only one, as well as 24 hubs–proteins with more than 30 partners–were detected. Previous studies in yeast have demonstrated that proteins acting as hubs are three times more likely to be essential for cells than proteins with only a small number of links (Jeong et al., 2001). Therefore, the hubs in our human interaction network deserve closer scrutiny with regard to important cellular tasks than other proteins.

To address the topological properties of our interaction map, we calculated the average clustering coefficient, C(k), a measure of the tendency of proteins in a network to form clusters or groups (Barabasi and Oltvai, 2004). We found that the average C(k) diminishes when the number of interactions per protein increases
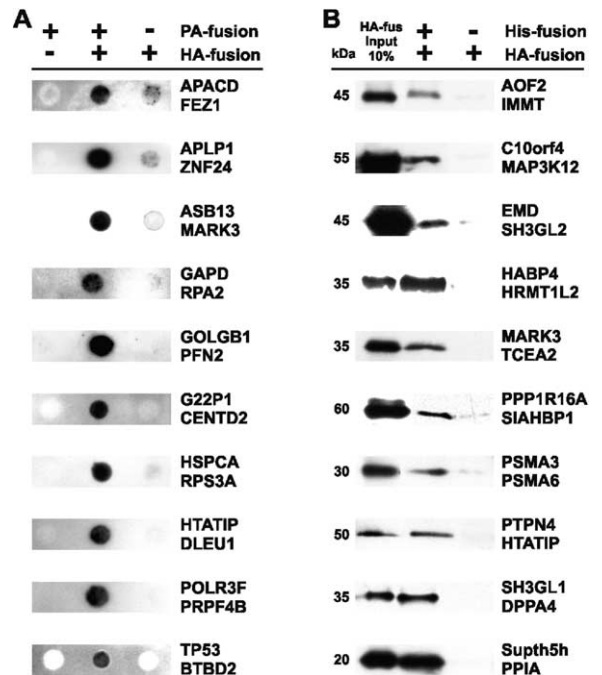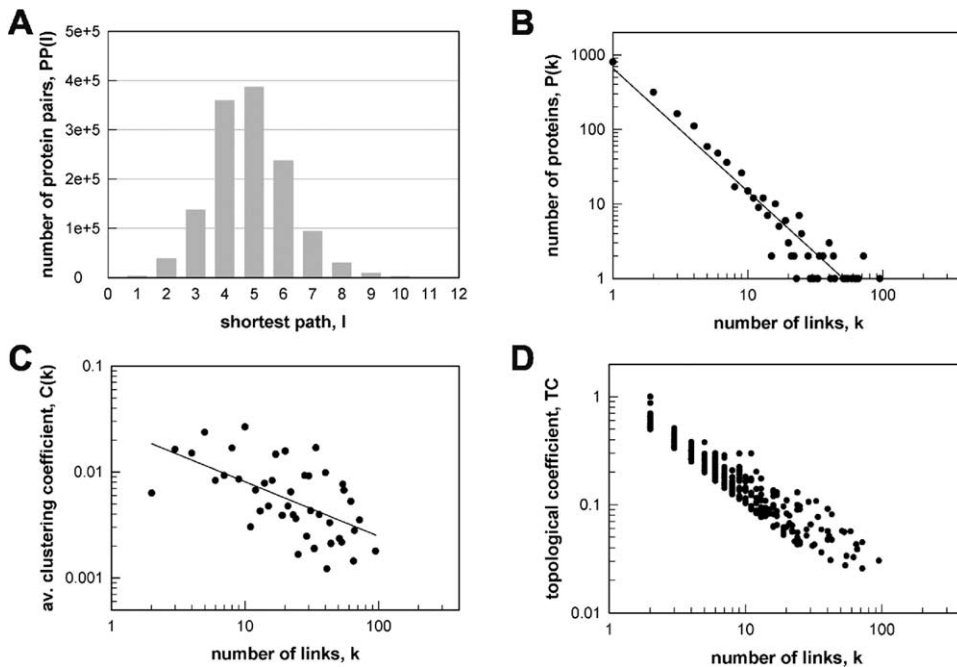


Figure 2. Verification of Y2H Interactions by Coimmunoprecipitation and Pull-Down Assays

(A) Membrane filter coimmunoprecipitation assay. Pairs of proteins were transiently expressed as hemagglutinin (HA)- and protein A (PA)-tagged fusions in COS-1 cells. Cleared cell lysates were filtered through a membrane coated with human IgG to retain the protein A fusion protein. After washing, the HA-tagged protein, bound to the protein A fusion partner, was detected on the membrane using anti-HA antibody. Identities of the PA- and HA-tagged fusions are as indicated. Protein expression levels and nonrelated control experiments are presented in Figure S1.

(B) In vitro pull-down assay. His-tagged fusions proteins produced in *E. coli* were immobilized, and their interacting partners (HA-fusions) were pulled down from COS-1 cell extracts. Binding was detected by SDS-PAGE and immunoblotting using anti-HA antibody. Identities of the His- and HA-tagged fusions are as indicated.

**Figure 3. Properties of the Human Y2H PPI Network**

(A) Distribution of the shortest path (l) between pairs of proteins in the Y2H network. On average, any two proteins in the network are connected via 4.85 links.

(B) Degree distribution of the network proteins. Number of proteins with a given link (k) in the network approximates a power-law ($P(k) \sim k^\gamma$; $\gamma = 1.78$).

(C) Degree distribution of the clustering coefficients of the network proteins. The average clustering coefficient of all nodes with k links was plotted against the number of links. ($CC_p = 2n/k_p(k_p - 1)$, with n as the number of links connecting the $k_p$ neighbors of node p to each other; see also Figure S3B).

(D) Degree distribution of the topological coefficients of the network proteins. The topological coefficient was calculated for every protein in the network and plotted against the number of links ($TC_p = average(J(p,j)/k_p)$, where J(p,j) denotes the number of nodes to which both p and j are linked, $k_p$ is the number of links of node p; see also Figure S3C).

(Figure 3C), indicating that the network has a potential hierarchical organization (Barabasi and Oltvai, 2004). A very similar result was also obtained when human interactions of a *HPRD* reference data set were analyzed (Figure S3B). In hierarchical networks, sparsely connected proteins are part of highly linked regions, which are connected via hubs (Ravasz et al., 2002). This also suggests that the interaction network has two levels of organization, local clustering, potentially representing protein complexes or functional modules, and more global connectivity mediated via hubs, conceivable as higher-order communication points between protein complexes (Han et al., 2004).

Besides the clustering coefficient, the topological coefficient was used (Goldberg and Roth, 2003; Ravasz et al., 2002) to study the characteristics of the interaction network. The topological coefficient, TC(k), is a relative measure for the extent to which a protein in the network shares interaction partners with other proteins. (see Figure S3C). As shown in Figure 3D, also the topological coefficient decreased with the number of links (close to 1/k), demonstrating that, relatively, in our network, hubs do not have more common neighbors than proteins with fewer links. This indicates that proteins with many links are not artificially clustered together (Supplemental Data and Table S2). Moreover, it confirms the modular network organization indicated by the clustering coefficient.

**Establishment of Criteria for the Development of a Confidence-Scoring System**

To enable a meaningful evaluation of the potential biological relevance of the identified interactions, it is critical to assess the confidence of an interaction (Formstecher et al., 2005; Giot et al., 2003; Li et al., 2004). For this purpose, criteria for confidence need to be defined. Using experimental, topological, and GO information, we established the following six criteria for confidence classification:

(1) The PPI activates three reporter genes (*HIS3*, *URA3*, and *lacZ*).
(2) The PPI is found in human interaction clusters.
(3) The PPI is found in orthologous *D. melanogaster* clusters.
(4) The PPI is found in orthologous *C. elegans* clusters.
(5) The PPI is found in orthologous *S. cerevisiae* clusters.
(6) The PPI is formed of two proteins sharing GO annotation.

***The Reporter Gene Activation Criterion***

Y2H interactions that can be identified by activation of three reporter genes are of higher confidence and can be reproduced with higher success rates than interac-

Table 1. Criteria for the Selection of Potential Higher-Confidence Interactions

| Confidence Criteria | Definition of the Criteria | Number of PPIs Selected |
| --- | --- | --- |
| Crit. 1 | *HIS3*, *URA3*, and *lacZ* Y2H reporter activity | 2054 |
| Crit. 2 | PPIs found in human interaction clusters | 1813 |
| Crit. 3 | PPIs found in orthologous *D. melanogaster* clusters | 957 |
| Crit. 4 | PPIs found in orthologous *C. elegans* clusters | 479 |
| Crit. 5 | PPIs found in orthologous *S. cerevisiae* clusters | 316 |
| Crit. 6 | PPIs formed of proteins sharing GO annotation | 130 |
| | Total selected interactions | 5749 |

tions detected only with two reporters. This is supported by our data (c.f. Figures 1 and 2) as well as several other high-throughput interaction-mapping studies (Vidalain et al., 2004). Therefore, the 2054 interactions (LacZ4 set) that were identified with the reporters *HIS3*, *URA3*, and *lacZ* were regarded as of higher confidence (Table 1, Crit. 1), while the 1145 interactions which were detected only with the *HIS3* and *URA3* growth reporters (SD4 set) were not.

### The Topological Criteria

*PPIs in Human Interaction Loop Motifs.* As cellular functions are carried out by stably or transiently associated groups of proteins, we reasoned that interactions in potential functional modules (Barabasi and Oltvai, 2004; Milo et al., 2002) are of higher confidence than others. Therefore, Y2H interactions that are present in three- and four-protein-interaction loop motifs (Goldberg and Roth, 2003; Wuchty et al., 2003; Yeger-Lotem et al., 2004) were identified. For this purpose, we combined our Y2H with the human PPI reference data extracted from *HPRD* (Peri et al., 2003) in order to create a denser, more comprehensive data set for motif analysis (for a detailed description of the interaction motif analysis see the Supplemental Data). Our in silico analysis revealed 1809 Y2H interactions that participate in three- and four-protein-interaction loops. Sixteen interactions from the *HPRD* data were directly recapitulated in our Y2H data (3%), which is in agreement with currently expected overlap rates between data sets (Formstecher et al., 2005; Han et al., 2005). Four of these did not participate in protein-interaction loops. In total, 1813 interactions were classified as of higher confidence based on topological criteria for human interactions (Table 1, Crit. 2).

*PPIs in Orthologous Interaction Loop Motifs.* In order to utilize the large interaction data sets of model organisms for the identification of higher confidence interactions by loop motif analysis, we determined which of the interacting proteins from our network have orthologous proteins in *D. melanogaster* (Dm), *C.elegans* (Ce), and *S. cerevisiae* (Sc) (Remm et al., 2001). Then, we used this information to assemble theoretical orthologous interactions that emulate the respective human Y2H PPIs. Next, we merged these data sets with published model organism PPI data from Dm (Giot et al., 2003), Ce (Li et al., 2004), and Sc (Mewes et al., 2004) and determined the appearance of orthologous Y2H interactions in interaction loop motifs. Using this in silico approach 957, 497, and 316 interactions were identified in the Dm, Ce, and Sc data sets, respectively, and regarded as of higher confidence (Table 1, Crit. 3, 4, and 5). Subsumed in these sets were 35 truly conserved in-

teractions–interologs (Matthews et al., 2001; Walhout et al., 2000).

The significance of the motif analyses for interaction confidence was evaluated by computer-generated randomized networks with the same properties as the Y2H network (scale-free, same degree distribution). We compared the number ($N_{Y2H}$) of interactions in loop motifs in the Y2H and the randomized networks; then, Z score values (Z score = ($N_{Y2H}$ − $N_{rand}$)/SD) were calculated as a qualitative measure of statistical significance (Yeger-Lotem et al., 2004). As shown in Figure 4A, the number of times Y2H interactions appear in human interaction loops is more than 25 standard deviations greater than their mean number of appearances in randomized networks ($N_{rand}$), indicating that the utilization of loop motifs as a confidence criterion is justified. Using randomized networks of orthologous proteins, similar results were obtained for Dm data. Lower Z score values were measured for Ce and Sc data, which might reflect a lower degree of homology to human protein interactions (Figure 4A).

### The GO Coannotation Criterion

As proteins of similar cellular function and localization tend to form PPI clusters, the Y2H PPIs were also investigated with regard to GO annotation (Lehner and Fraser, 2004). We examined how many Y2H interaction partners appear together on the same GO hierarchy level and compared these numbers with the analogous numbers for the randomized networks. As shown in Figure 4B, from a depth of seven onward in the GO hierarchy, significantly higher numbers of Y2H interactions with shared terms in all three GO classes (cellular component, biological process, and molecular function) were detected. As GO coannotated interaction pairs are generally regarded as more reliable (Giot et al., 2003; Ramani et al., 2005), these interactions (130 PPIs) were considered as of higher confidence (Table 1, Crit. 6).

### Identification of High-Confidence Interactions

To classify Y2H interactions into categories of low, medium, and high confidence, an interaction was awarded one quality point for each fulfilled criterion described above. 5749 points were given to 2618 of the 3186 PPIs, with any given interaction receiving between 0 and 6 quality points. Interactions were then ranked according to their number of quality points and grouped into the three confidence sets (Figure 4C). Five hundred sixty-eight (18%) Y2H interactions did not receive quality points and were classified as of low confidence (LC set). 1707 interactions (54%) obtained 1–2 quality points and were classified as of medium confidence (MC set).
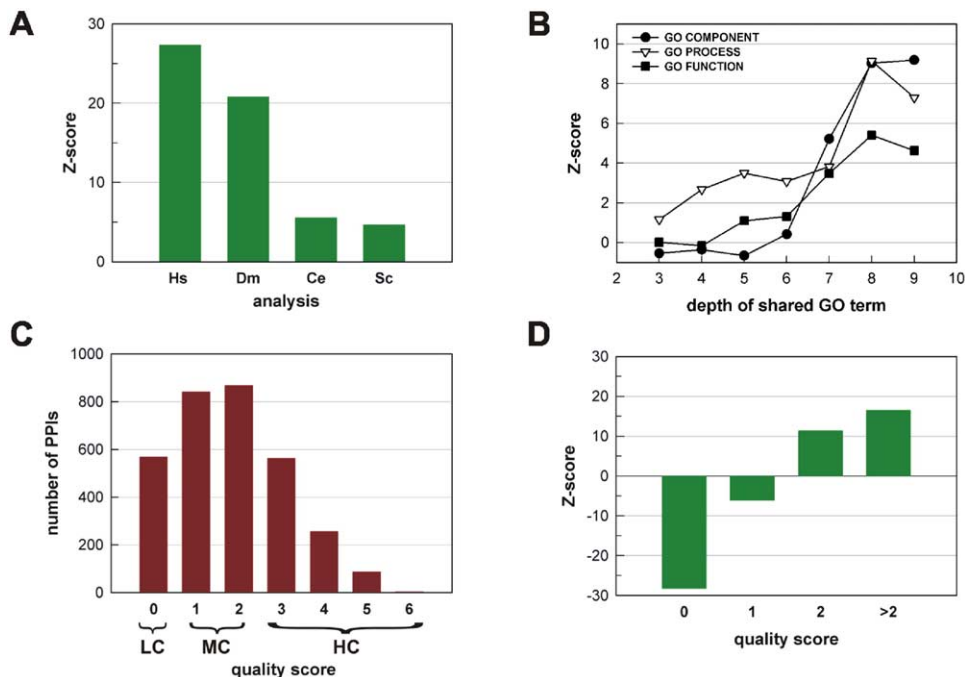
Figure 4. A Confidence-Scoring System for Y2H PPIs

(A) Statistical significance of loop motif analysis (criteria 2–5). Z score values (($\#PPIs_{Y2H} - \#PPIs_{rand}$)/$SD_{rand}$) were calculated from the number of loop motif interactions from the experimental Y2H network and the random model networks (Table S6). Values for human (Hs), *Drosophila* (Dm), *C. elegans* (Ce), and *S. cerevisiae* (Sc) are presented. Y2H interactions appear with higher frequency in protein interaction loops than interactions from the randomized networks.

(B) Statistical significance of interacting proteins with shared GO annotation at various levels of the GO hierarchy for component, process, and function. Z score values (($\#PPIs_{Y2H} - \#PPIs_{rand}$)/$SD_{rand}$) were calculated from the number of interacting proteins with shared GO terms from the experimental Y2H and random networks (Table S7). Interacting proteins annotating to the same GO categories at a depth of $\geq 7$ fulfilled criterion 6.

(C) Distribution of the Y2H interactions according to their quality scores. Based on experimental and bioinformatic criteria (criteria 1–6), 5749 quality points were awarded to 2618 Y2H PPIs. Low (LC)-, medium (MC)-, and high (HC)-confidence data sets are indicated.

(D) Statistical significance of the quality scoring system. Z score values (($\#PPIs_{Y2H} - \#PPIs_{rand}$)/$SD_{rand}$) were calculated from the number interactions that scored quality points in the bioinformatic analyses from the Y2H and random networks (Table S8).

Importantly, the data allowed the definition of 911 high-confidence interactions that collected 3 or more quality points (HC set, 28%), involving 401 different human proteins.

Finally, we determined whether the assignment of Y2H interactions into confidence sets is statistically significant. We quantified the numbers of Y2H interactions in the confidence sets and compared the results with the numbers obtained for interactions from randomized networks. As shown in Figure 4D, Y2H interactions were found with higher frequency in the HC set (3 or more quality points) than interactions from random networks. This indicates that our scoring system for the selection of interactions with potential higher biological relevance yields statistically significant results.

**Properties of the High-Confidence PPI Network**

The group of HC interactions resulting from our confidence-scoring procedure mainly contains interactions with biological context information, i.e., interactions where additional supportive information, like GO coannotation or participation in loop motifs, is available. Therefore, we suggest that the HC interactions are

most promising with regard to further functional analysis and the generation of new hypotheses. Because of this increased relevance, we specially extracted a network of the HC interactions from our data, which is charted in Figure 5A. We grouped the proteins in three broad categories using GO and OMIM criteria: disease proteins (45), uncharacterized proteins (49), and known proteins (307). For the 45 disease proteins, 163 HC PPIs were identified (Table S1). This information can be used as a resource for disease-specific investigations.

As visible in Figure 5A, the majority of proteins (87%) are linked and form a large interaction network. In addition, 24 small networks with less than six proteins were obtained. The proteins contained in the HC set are still a nonbiased representation of the human proteome when GO criteria are applied (Figure S5), indicating that confidence filtering did not preferentially remove certain groups/classes of proteins. In contrast to other confidence-scoring procedures (Formstecher et al., 2005; Giot et al., 2003), our approach did not exclude proteins because they have a high number of links. Therefore, the HC and the complete data set have a very similar degree distribution and relative number of hubs. However, further topological analysis revealed
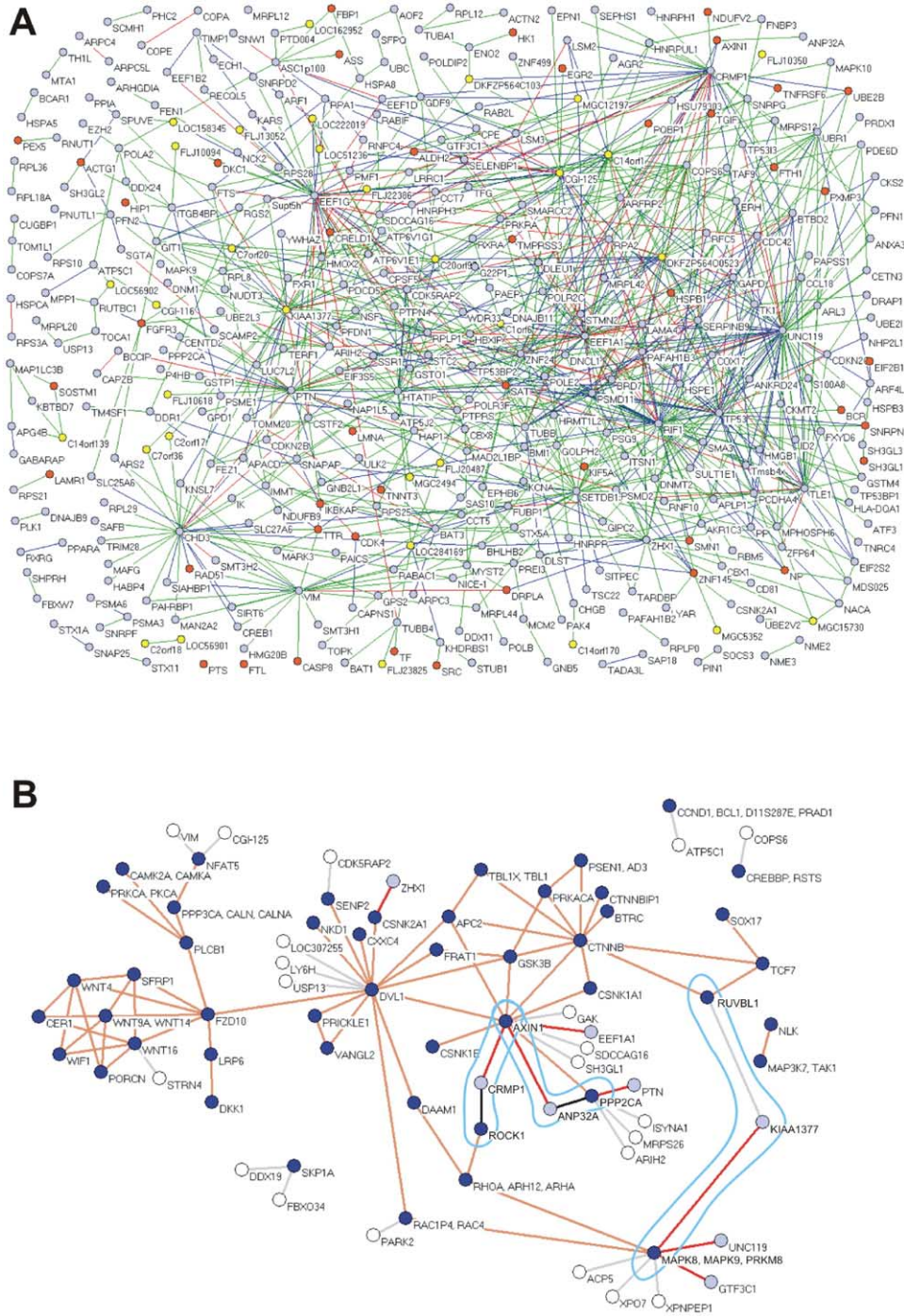
**Figure 5. Network Views**

(A) A graph of the HC interaction network involving 401 proteins linked via 911 interactions. Orange: disease proteins (according to OMIM morbidmap, NCBI); light blue: proteins with GO annotation; yellow: proteins without GO and disease annotation. Interactions connecting the nodes are represented by color-coded lines according to their confidence scores. Green: 3 quality points; blue: 4 quality points; red: 5 quality points; purple: 6 quality points.

(B) Y2H proteins linked to the Wnt signaling pathway. Spheres: proteins annotated in the KEGG regulatory pathways (blue); proteins with high-confidence Y2H interactions (light blue); proteins with MC or LC PPIs (white). Links: protein-protein relations annotated in the Wnt pathway (brown); PPIs from HPRD (black); Y2H HC PPIs (red); Y2H LC and MC PPIs (gray). The proteins CRMP1, ANP32A, and KIAA1377 bridge two proteins in the Wnt pathway.

that the mean path length between proteins is significantly shorter and the average clustering coefficient is higher in the HC map (Figure S5). These results suggest that the HC network has an even more pronounced hierarchical structure than the complete network and contains a larger number of closely linked protein clusters, which could be indicative of functional protein complexes (Barabasi and Oltvai, 2004).

### Development of a Database for Network Exploration

Accompanying this manuscript, a web-based searchable database was created at http://www.mdc-berlin. de/neuroprot/database.htm. This database permits queries for protein names and synonyms, accession numbers, gene names, and official gene symbols, as well as LocusLinkID. Annotations are provided for every protein and every interaction it is involved in. The database also enables a graphical representation of queried proteins and their interaction partners. Links to bibliographic references and relevant external databases are also included. A screen-shot illustration of the user interface is presented in the Supplemental Data.

### Linking Y2H Interactions to Regulatory Pathways

A direct comparison between the Y2H network and 22 human regulatory pathways from the Kyoto encyclopedia of genes and genomes—KEGG (Kanehisa et al., 2004)—revealed that the two data sets share 162 proteins involved in different signal transduction processes or neurodegenerative diseases (see Table S4 for a survey). We found that these proteins had 568 interactions in the Y2H network, 168 of which were HC. They link 115 proteins, including 13 disease proteins, to the 22 different regulatory pathways. Eight proteins, e.g., (ZHX1, PTN, EEF1A1, ANP32A, CRMP1, GTF3C1, UNC119, and KIAA1377) were found to connect via HC interactions with proteins in the Wnt signaling pathway (Figure 5B), which controls patterning and organogenesis during development and is important for tumor formation in adults (Logan and Nusse, 2004).

Using bioinformatic tools, we also identified proteins that form links with two or more proteins annotated in a KEGG pathway, increasing the confidence of the pathway assignment. This approach identified 66 proteins, including seven disease proteins (Table S5). We found that the proteins ANP32A, CRMP1, and KIAA1377, e.g., are linked to the Wnt pathway via two proteins (Figure 5B). Interestingly, ANP32A (*acidic leucine-rich nuclear phosphoprotein 32*), a potential tumor suppressor (Bai et al., 2001), associates with Axin-1 (Luo and Lin, 2004), as well as phosphatase 2A (PPP2CA), which are both crucial for regulating the Wnt pathway. In addition, an interaction between Axin-1, ROCK1, and CRMP1 (collapsin response mediator protein-1) was identified. ROCK1 is part of the noncanonical Wnt/PCP pathway and modulates cytoskeletal dynamics and the activity of MAP kinases, while CRMP1 functions in Rho/Rac signaling during neuronal differentiation (Arimura et al., 2004), suggesting a new link between Wnt signaling and processes controlling cytoskelatal organization. Finally, interactions linking RUVBL1, a β-catenin/Tcf cofactor (Bauer et al., 2000), to MAPK9, a nuclear member of the MAP kinase family (Luo and Lin, 2004),

via the uncharacterized protein KIAA1377 were also found. This suggests that KIAA1377 could participate in the integration/diversification of signals at transcription level.

### ANP32A and CRMP1 Modulate the Activity of the Wnt Pathway

In order to determine the binding sites required for the interactions between ANP32A, CRMP1, and Axin-1, GST pull-down experiments were performed (Figure 6). We found that the acidic C-terminal half of ANP32A (aa 150–249), which contains a potential acetyltransferase inhibitory domain (Seo et al., 2002), interacts with the C-terminal DIX domain of Axin-1 (aa 725–826), while the N-terminal half, with three leucine-rich repeats (Ulitzur et al., 1997), did not bind to Axin-1. Also, we showed that the C-terminal 228 amino acids of CRMP1 are critical for the interaction with a central region in Axin-1 (aa 510–625; Figures 6A and 6B).

To test whether the interactions are functionally relevant in the canonical Wnt signaling cascade, a cell-based transcription assay was performed. Axin-1 is a negative regulator of the Wnt pathway that controls the levels of the transcriptional activator β-catenin (Behrens et al., 1996). In the absence of Wnt signals, β-catenin is degraded by an Axin-1-containing multiprotein complex (Luo and Lin, 2004), while activation of the pathway with, e.g., dishevelled (Dvl) inhibits β-catenin degradation and induces Lef/Tcf-dependent transcription (Huelsken and Birchmeier, 2001). In the assay, moderate overexpression of Dvl in HEK293 cells caused a 4-fold increase in reporter gene activity. However, this increase was reduced to 1.5-fold in a dose-dependent manner, when full-length ANP32A or CRMP1 were coexpressed (Figure 6C). The activity of protein fragments in the Wnt signaling assay correlated with their ability to bind Axin-1. The C-terminal half of ANP32A containing the Axin-1 binding site was sufficient to suppress Wnt signaling, while the N-terminal 150 amino acids of ANP32A were inactive. Similarly, the C-terminal fragment of CRMP1 containing the Axin-1 binding site reduced Lef/Tcf-dependent transcription as efficiently as the full-length protein, while the N-terminal part had no effect (Figure 6C). These results demonstrate that connecting proteins via Y2H interactions to signaling cascades such as the Wnt pathway allows the identification of potential pathway modulators.

### Discussion

To assign functions to uncharacterized proteins and to understand the composition of protein complexes, several large- and medium-scale interaction studies have been undertaken using the Y2H system or MS-based functional proteomics approaches (von Mering et al., 2002). These studies have provided the scientific community with predictions on protein function of model organisms such as yeast, *Drosophila*, or *C. elegans*. However, proteome-wide maps of interactions among human proteins have not yet been presented.

Here, we report a first systematic Y2H analysis of human proteins. Using a matrix approach and two rounds of automated interaction mating, we identified 3186 in-
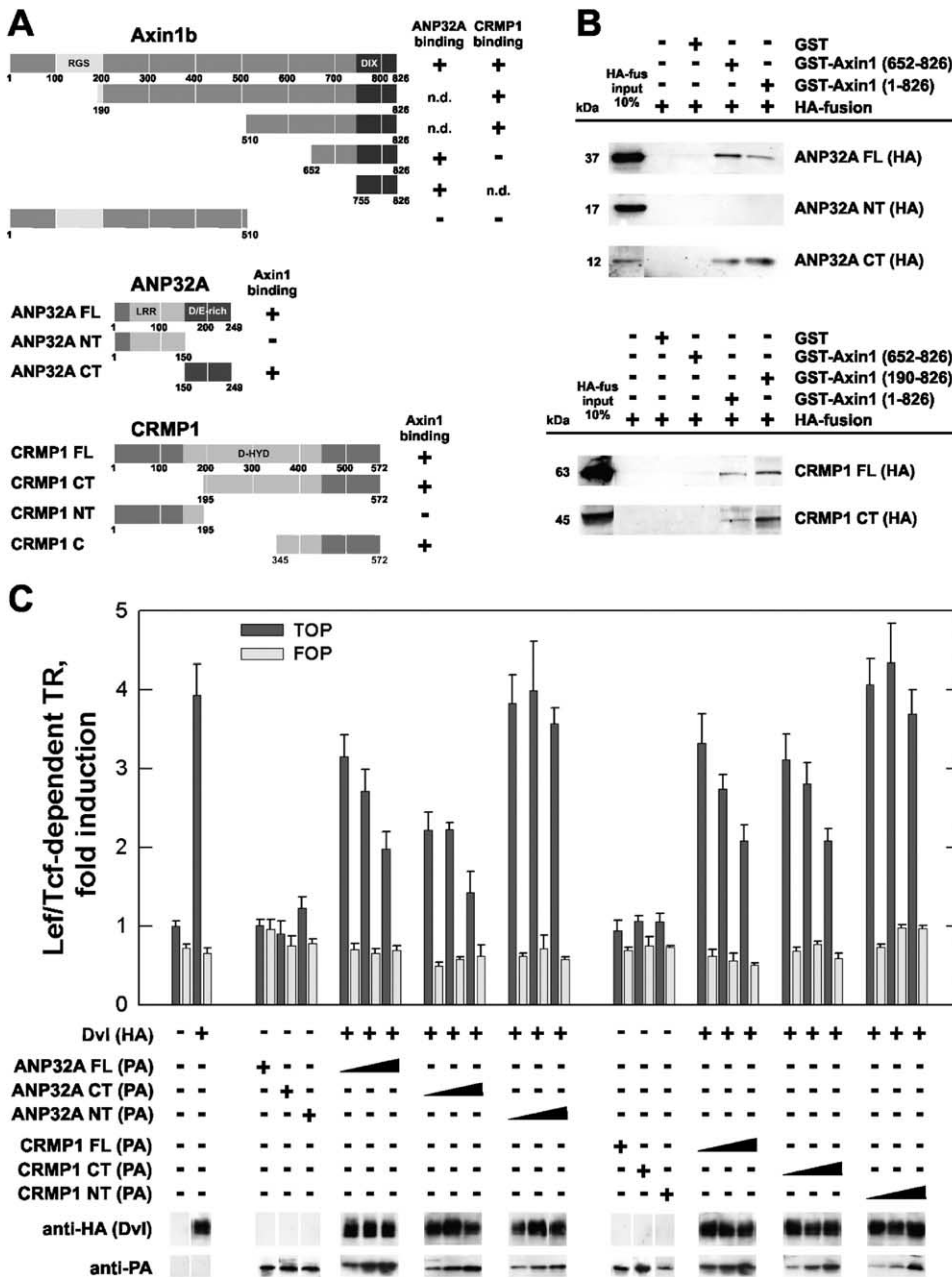
Figure 6. The Axin-1 Interacting Proteins ANP32A and CRMP1 Act as Repressors in Wnt/β-catenin Signaling

(A) Schematic representation of the primary structures of Axin-1, ANP32A, and CRMP1. RGS: regulation of G protein signaling domain; DIX: dishevelled and axin domain; LRR: leucine-rich repeat; D/E-rich: Asp/Glu-rich region; D-HYD: similarity to dihydropyrimidinases. Binding results are summarized.

(B) GST pull-down assays to confirm the Axin-1/ANP32A (top) and the Axin-1/CRMP1 interactions.

(C) ANP32A and CRMP1 repress Lef/Tcf-dependent transcription induced by Dishevelled (Dvl) in cell-based assays. TOP reporter: gray bars; FOP reporter: white bars. The averages from at least three assays for every construct were combined. Expression levels of the proteins were verified by Western blotting using anti-HA antibody recognizing Dvl and anti-PA antibody detecting ANP32A and CRMP1 proteins.

teractions connecting 1705 proteins. Among them, 195 disease proteins and 342 uncharacterized proteins were placed in a new context via direct and indirect interactions with other proteins. We validated the interactions by independent pull-down and coimmunoprecipitation experiments (Figure 2), as interactions detectable in different binding assays are unlikely to be experimental false positives (Goehler et al., 2004; Li et al., 2004). The overall success rate was about 65%, confirming our screening procedure as capable of generating a data set that contains a large fraction of reliable interactions.

Besides experimental false positives arising from the inherent limitations of the Y2H approach, biological

false positives reduce the reliability of large-scale Y2H interaction data sets. Biological false positives are interactions that are produced and reproduced in various exogenous assays but do not form under physiological conditions. Due to the scarcity of information about stable and transient protein complexes in human, it seems impossible to estimate the rate of biological false positives in our data set. Their frequency cannot be determined without additional information from other proteomics studies providing physiological data about protein complex composition, protein localization, and cell type specificity (Ge et al., 2003).

High rates of missed interactions have been reported when different Y2H screens or data sets generated with different methods were crossvalidated (Formstecher et al., 2005; von Mering et al., 2002). The studies cover only small fractions of the interactome, and approaches differ widely enough to result in complementary rather than overlapping data. Various explanations for the occurrence of false negatives in the Y2H approach have been put forward, mostly relating to the lack of post-translational modifications, improper folding of the hybrid proteins, or the inability of interacting proteins to enter the nucleus (von Mering et al., 2002). We evaluated the overlap of Y2H interactions with previously published data sets like *HPRD*, which collects interactions identified in small-scale experiments with various in vitro and in vivo techniques, finding a similar lack of concordance. This might have to do with the mentioned insufficiencies of the Y2H system or the patchiness of the information of the reference data sets (Ramani et al., 2005). However, Han et al., (2005) recently attempted to explain that the low coverage of interaction data sets is to be expected without assuming false positives, because currently available methods are only capable of producing incomplete data sets. We suggest that novel identification as well as multiple-step validation/confirmation strategies will have to be put in place in order to produce comprehensive, crossconnected interaction data sets.

To identify interactions that are biologically meaningful, a confidence-scoring system was developed using experimental, topological, and GO criteria (Figure 4). Protein clusters were detected significantly more often in the Y2H network than in any of the random control networks, indicating that biologically relevant, functional complexes can be recognized in the interaction data (Supplemental Data). For example, we identified new interaction partners for the disease protein emerin (EMD), which causes X-linked Emery-Dreifuss muscular dystrophy (EDMD) when mutated (Emery, 2002). Emerin binds to the Src-homology 3 proteins SH3GL2 and SH3GL3 and the uncharacterized developmental pluripotency-associated protein 4 (DPPA4), which itself binds to SH3GL3 and SH3GL1, thus forming a small, highly connected interaction cluster. The identification of these interactions might promote understanding of EMD function in mammalian cells and can be used as a starting point for investigating the role of the participating proteins in EDMD pathogenesis.

Another interesting outcome of our study is the high connectedness of network proteins to proteins of gene regulatory pathways listed in the KEGG database. Using simple bioinformatic tools, we directly linked about 150 human proteins via HC interactions to the 22 different KEGG pathways. In addition, proteins that bridge two or more proteins in a given pathway were identified. Utilizing this more stringent approach, 66 proteins could be newly mapped to one or more KEGG pathways (Tables S4 and S5). Using cell-based assays, we functionally validated two of the proteins, which were linked to the Wnt pathway by computational analysis (Figure 6). We found that the proteins ANP32A and CRMP1, which are both implicated in disease processes (Bai et al., 2001; Shih et al., 2003), can suppress canonical Wnt signaling, suggesting that they might be functional modulators of this pathway in vivo. This indicates that Y2H screening combined with bioinformatic pathway mapping increases our knowledge about signaling cascades and permits design of new experimental strategies.

## Conclusions

This report supplies more than 3000 protein-protein interactions, 911 supported with topological and GO criteria, 159 verified biochemically. The interaction map links 195 disease proteins to previously unidentified partners, allows the description of 342 uncharacterized human proteins via their interactions, and suggests new roles for hundreds of known proteins. Also, the study integrates Y2H interaction data with known regulatory pathways, extracting potential functional modules that participate in signaling cascades from the static, large-scale human Y2H map. This human PPI map will serve as a unique resource for further experimentation and analysis leading to the identification of disease-modifier genes and new drug targets.

## Experimental Procedures

### Subcloning of Human cDNAs into Y2H Plasmids

9665 cDNAs of the hEx1 library (Bussow et al., 1998) were sequenced from the 5′ end to determine the identity and the reading frame of each cDNA fragment. BLASTP analysis against the nr (NCBI) or TrEMBL (Swiss-Prot) databases revealed a nonredundant set of 4275 cDNAs, which were inserted by restriction cloning into the Y2H plasmids pGAD426 and pBTM117c (http://www.mdc-berlin. de/neuroprot/labequip.htm). For recombinational cloning of human full-length ORFs, 2136 human cDNA fragments were PCR amplified with specific primer pairs from source clones of the RZPD repository and BP cloned into pDONR201 (Invitrogen, Carlsbad). Recombinant clones were sequenced and annotated using BLASTP searches. cDNA fragments were then shuttled into the Y2H vectors pBTM116-D9 and pGAD426-D3 by recombinational cloning (http:// www.rzpd.de/products/orfclones/). The redundancy of clones in the matrix, i.e., clones coding for different parts of the same proteins, was less than 4%. In total, 48% of the plasmids encoded full-length ORFs, whereas 52% coded for C-terminal fragments of larger human proteins.

### Automated Y2H Screening

To create a matrix for interaction mating, the L40ccα *MAT*α yeast strain (Goehler et al., 2004) was individually transformed with prey plasmids (coding Gal4 activation domain fusions); the resulting yeast clones were arrayed in 384-well microtiter plates. Simultaneously, the bait plasmids (coding LexA DNA binding domain fusions) were introduced into a L40ccU *MAT*a strain and assembled in 96-well plates. Baits (19.6%), which activated the *HIS3*, *URA3*, and *lacZ* reporter genes after mating with a *MAT*α strain expressing an AD protein, were excluded from the automated Y2H analysis.

For interaction mating, 5 μl liquid cultures of the *MAT*α yeast strains were replicated in 384-well MTPs using a pipetting robot

(Biomek FX), grown, and mixed with 40 μl pooled *MAT*a strains (eight baits). The yeast mixtures were then transferred onto YPD agar plates using a spotting robot (KBiosystems) and incubated for 36 hr at 30°C. After mating, the clones were automatically picked from the plates and transferred into 384-well MTPs containing SDII (-Leu-Trp) liquid medium. For selection of PPIs, diploid yeasts were spotted onto SDIV (-Leu-Trp-Ura-His) agar plates as well as nylon membranes placed on SDIV agar plates. After 5–6 days of incubation at 30°C, digitized images of the agar plates and nylon membranes were assessed for growth and β-galactosidase activity using the software Visual Grid (GPC Biotech).

For confirmation of interactions, the eight baits from each pool were arrayed in 96-well MTPs (Biomek 2000) and mated with the positive preys identified in the first mating screen. After 36 hr at 30°C, yeast cultures were spotted onto SDII agar plates for selection of diploid cells expressing both protein fusions. After 4 days at 30°C, the yeast colonies were assayed on SDIV agar plates and nylon membranes.

### Membrane Coimmunopurification and Pull-Down Assays

For expression of hemagglutinin (HA)- and protein A (PA)-tagged fusions, cDNA fragments (identical to those in the Y2H assays) were subcloned into pTL-HA or pcDNA3.1-PA, respectively, and cotransfected pairwise into COS-1 cells. Cell extracts were assessed for the expression of both proteins by SDS-PAGE and immunoblotting and filtered (approximately 3 μg protein extract) through a nitrocellulose membrane (Schleicher & Schuell) coated with human IgG (Sigma, 1:1000 in PBS) using a 96-well dot blot apparatus. Membranes were washed six times and probed with the anti-HA monoclonal antibody 12CA5 (Roche Diagnostics) for detection of PPIs.

For in vitro pull-down assays, cDNA fragments were subcloned into pQE30-NST (Bussow et al., 1998) or pTL-HA. Soluble protein extracts of His- and HA-tagged fusions were prepared from *E. coli* and COS-1 cells, respectively. His-tagged fusions were bound to $Ni^{2+}$NTA agarose beads (approximately 30 μg protein). After washing (50 mM HEPES-KOH [pH 7.4], 300 mM NaCl, 5 mM imidazol, 1 mM DTT), they were incubated for 2 hr at 4°C with 200 μg COS-1 cell extract containing the potential interacting HA-tagged fusion protein. After washing, bound proteins were analyzed by SDS-PAGE and immunoblotting with the anti-HA antibody.

### Luciferase Reporter Assay

Lef/Tcf reporter assays were performed as previously described (Brembeck et al., 2004) with the following modifications: HEK293 cells were cotransfected with 0.125, 0.25, and 0.5 μg of the indicated ANP32A or CRMP1 plasmids and 0.5 μg Dvl expression constructs together with TOP or control FOP luciferase reporter and β-galactosidase plasmids using Lipofectamine 2000 (Invitrogen). Empty vector DNA was added to a total of 1.25 μg plasmid DNA. Luciferase activity was determined 48 hr after transfection and normalized against β-galactosidase activity. Reporter assays were performed as triple transfections.

### Computational Analysis of the PPI Map

Using LL.out_hs (NCBI, 05-13-2004 release), proteins were mapped to a unique gene locus via their accession numbers obtained in a BLASTP search against the nr (NCBI) or TrEMBL (Swiss-Prot) databases. The BLAST analysis of Y2H clones and the orthology assignments computed with the InParanoid program (Remm et al., 2001) against the predicted proteome of *D. melanogaster* (FlyBase release r3.2.0), *C. elegans* (WormBase release WS121), and *S. cerevisiae* (SGD ORF set of 04-02-2004) are presented in Table S2. A PPI human reference set of 14,384 PPIs between 4,478 proteins was obtained from the *HPRD* (status 09-17-04). For the assignment of interolog clusters, we referred to the complete *D. melanogaster* Y2H data set (Giot et al., 2003; 20,439 PPIs; 6,991 proteins), the *C. elegans* WI5 data set (Li et al., 2004; 5534 PPIs; 3227 proteins), and the manually curated catalog of PPIs from *S. cerevisiae* (MIPS at http://mips.gsf.de/; 8946 PPIs; 4525 proteins). Topological analysis, e.g., degree distribution, clustering coefficients, and path lengths, was carried out with TopNet (Yu et al., 2004). GO assignments made use of NCBI loc2go, (12-09-2004) and OBO (12-09-

2004), and pathway assignment was performed using KEGG data (Release 34.0).

### Supplemental Data

Supplemental Data include five figures, supplemental text, and eight tables which provide a full documentation of the human PPI dataset and can be found with this article online at http://www.cell.com/cgi/content/full/122/6/957/DC1/.

### References

Arimura, N., Menager, C., Fukata, Y., and Kaibuchi, K. (2004). Role of CRMP-2 in neuronal polarity. J. Neurobiol. *58*, 34–47.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29.

Bai, J., Brody, J.R., Kadkol, S.S., and Pasternack, G.R. (2001). Tumor suppression and potentiation by manipulation of pp32 expression. Oncogene *20*, 2153–2160.

Barabasi, A.L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. Nat. Rev. Genet. *5*, 101–113.

Bauer, A., Chauvet, S., Huber, O., Usseglio, F., Rothbacher, U., Aragnol, D., Kemler, R., and Pradel, J. (2000). Pontin52 and reptin52 function as antagonistic regulators of beta-catenin signalling activity. EMBO J. *19*, 6121–6130.

Behrens, J., von Kries, J.P., Kuhl, M., Bruhn, L., Wedlich, D., Grosschedl, R., and Birchmeier, W. (1996). Functional interaction of beta-catenin with the transcription factor LEF-1. Nature *382*, 638–642.

Bork, P., Jensen, L.J., von Mering, C., Ramani, A.K., Lee, I., and Marcotte, E.M. (2004). Protein interaction networks from yeast to human. Curr. Opin. Struct. Biol. *14*, 292–299.

Brembeck, F.H., Schwarz-Romond, T., Bakkers, J., Wilhelm, S., Hammerschmidt, M., and Birchmeier, W. (2004). Essential role of BCL9-2 in the switch between beta-catenin's adhesive and transcriptional functions. Genes Dev. *18*, 2225–2230.

Bussow, K., Cahill, D., Nietfeld, W., Bancroft, D., Scherzinger, E., Lehrach, H., and Walter, G. (1998). A method for global protein expression and antibody screening on high-density filters of an arrayed cDNA library. Nucleic Acids Res. *26*, 5007–5008.

Colland, F., Jacq, X., Trouplin, V., Mougin, C., Groizeleau, C., Hamburger, A., Meil, A., Wojcik, J., Legrain, P., and Gauthier, J.M. (2004). Functional proteomics mapping of a human signaling pathway. Genome Res. *14*, 1324–1332.

Emery, A.E. (2002). The muscular dystrophies. Lancet *359*, 687–695.

Formstecher, E., Aresta, S., Collura, V., Hamburger, A., Meil, A., Trehin, A., Reverdy, C., Betin, V., Maire, S., Brun, C., et al. (2005). Protein interaction mapping: a Drosophila case study. Genome Res. *15*, 376–384.

Ge, H., Walhout, A.J., and Vidal, M. (2003). Integrating 'omic' information: a bridge between genomics and systems biology. Trends Genet. *19*, 551–560.

Giot, L., Bader, J.S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y.,

Hao, Y.L., Ooi, C.E., Godwin, B., Vitols, E., et al. (2003). A protein interaction map of *Drosophila melanogaster*. Science *302*, 1727–1736.

Goehler, H., Lalowski, M., Stelzl, U., Waelter, S., Stroedicke, M., Worm, U., Droege, A., Linderberg, K.S., Knoblich, M., Haenig, C., et al. (2004). A protein interaction network links GIT1, an enhancer of huntingtin aggregation, to Huntington's disease. Mol. Cell *15*, 853–865.

Goldberg, D.S., and Roth, F.P. (2003). Assessing experimentally derived interactions in a small world. Proc. Natl. Acad. Sci. USA *100*, 4372–4376.

Han, J.D., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J., Cusick, M.E., Roth, F.P., and Vidal, M. (2004). Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature *430*, 88–93.

Han, J.D., Dupuy, D., Bertin, N., Cusick, M.E., and Vidal, M. (2005). Effect of sampling on topology predictions of protein-protein interaction networks. Nat. Biotechnol. *23*, 839–844.

Huelsken, J., and Birchmeier, W. (2001). New aspects of Wnt signaling pathways in higher vertebrates. Curr. Opin. Genet. Dev. *11*, 547–553.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. USA *98*, 4569–4574.

Jeong, H., Mason, S.P., Barabasi, A.L., and Oltvai, Z.N. (2001). Lethality and centrality in protein networks. Nature *411*, 41–42.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. Nucleic Acids Res. *32*, D277–D280.

Lehner, B., and Fraser, A.G. (2004). A first-draft human protein-interaction map. Genome Biol. *5*, R63.

Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.O., Han, J.D., Chesneau, A., Hao, T., et al. (2004). A map of the interactome network of the metazoan *C. elegans*. Science *303*, 540–543.

Logan, C.Y., and Nusse, R. (2004). The Wnt signaling pathway in development and disease. Annu. Rev. Cell Dev. Biol. *20*, 781–810.

Luo, W., and Lin, S.C. (2004). Axin: a master scaffold for multiple signaling pathways. Neurosignals *13*, 99–113.

Matthews, L.R., Vaglio, P., Reboul, J., Ge, H., Davis, B.P., Garrels, J., Vincent, S., and Vidal, M. (2001). Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs". Genome Res. *11*, 2120–2126.

Mewes, H.W., Amid, C., Arnold, R., Frishman, D., Guldener, U., Mannhaupt, G., Munsterkotter, M., Pagel, P., Strack, N., Stumpflen, V., et al. (2004). MIPS: analysis and annotation of proteins from whole genomes. Nucleic Acids Res. *32*, D41–D44.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. Science *298*, 824–827.

Peri, S., Navarro, J.D., Amanchy, R., Kristiansen, T.Z., Jonnalagadda, C.K., Surendranath, V., Niranjan, V., Muthusamy, B., Gandhi, T.K., Gronborg, M., et al. (2003). Development of human protein reference database as an initial platform for approaching systems biology in humans. Genome Res. *13*, 2363–2371.

Rain, J.C., Selig, L., De Reuse, H., Battaglia, V., Reverdy, C., Simon, S., Lenzen, G., Petel, F., Wojcik, J., Schachter, V., et al. (2001). The protein-protein interaction map of *Helicobacter pylori*. Nature *409*, 211–215.

Ramani, A.K., Bunescu, R.C., Mooney, R.J., and Marcotte, E.M. (2005). Consolidating the set of known human protein-protein interactions in preparation for large-scale mapping of the human interactome. Genome Biol. *6*, R40.

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., and Barabasi, A.L. (2002). Hierarchical organization of modularity in metabolic networks. Science *297*, 1551–1555.

Remm, M., Storm, C.E., and Sonnhammer, E.L. (2001). Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J. Mol. Biol. *314*, 1041–1052.

Seo, S.B., Macfarlan, T., McNamara, P., Hong, R., Mukai, Y., Heo, S., and Chakravarti, D. (2002). Regulation of histone acetylation and transcription by nuclear protein pp32, a subunit of the INHAT complex. J. Biol. Chem. *277*, 14005–14010.

Shih, J.Y., Lee, Y.C., Yang, S.C., Hong, T.M., Huang, C.Y., and Yang, P.C. (2003). Collapsin response mediator protein-1: a novel invasion-suppressor gene. Clin. Exp. Metastasis *20*, 69–76.

Strogatz, S.H. (2001). Exploring complex networks. Nature *410*, 268–276.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., et al. (2000). A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature *403*, 623–627.

Ulitzur, N., Rancano, C., and Pfeffer, S.R. (1997). Biochemical characterization of mapmodulin, a protein that binds microtubule-associated proteins. J. Biol. Chem. *272*, 30577–30582.

Vidalain, P.O., Boxem, M., Ge, H., Li, S., and Vidal, M. (2004). Increasing specificity in high-throughput yeast two-hybrid experiments. Methods *32*, 363–370.

von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., and Bork, P. (2002). Comparative assessment of large-scale data sets of protein-protein interactions. Nature *417*, 399–403.

Walhout, A., Sordella, R., Lu, X., Hartley, J., Temple, G., Brasch, M., Thierry-Mieg, N., and Vidal, M. (2000). Protein interaction mapping in C. elegans using proteins involved in vulval development. Science *287*, 116–122.

Wuchty, S., Oltvai, Z.N., and Barabasi, A.L. (2003). Evolutionary conservation of motif constituents in the yeast protein interaction network. Nat. Genet. *35*, 176–179.

Yeger-Lotem, E., Sattath, S., Kashtan, N., Itzkovitz, S., Milo, R., Pinter, R.Y., Alon, U., and Margalit, H. (2004). Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction. Proc. Natl. Acad. Sci. USA *101*, 5934–5939.

Yu, H., Zhu, X., Greenbaum, D., Karro, J., and Gerstein, M. (2004). TopNet: a tool for comparing biological sub-networks, correlating protein properties with topological statistics. Nucleic Acids Res. *32*, 328–337.